

DOCUMENT RESUME

ED 128 535

UD 016 361

TITLE NAACP Report on Minority Testing.
 INSTITUTION National Association for the Advancement of Colored People, New York, N.Y.
 SPONS AGENCY College Entrance Examination Board, New York, N.Y.; Educational Testing Service, Princeton, N.J.
 PUB DATE May 76
 NOTE 43p.

EDRS PRICE MF-\$0.83 HC-\$2.06 Plus Postage.
 DESCRIPTORS *Achievement Tests; *Conference Reports; Criteria Referenced Tests; Culture Free Tests; Educationally Disadvantaged; *Educational Testing; Ethnic Groups; Intelligence Tests; Mental Tests; *Minority Group Children; Minority Groups; Norm Referenced Tests; *Student Testing; Test Bias; Test Construction; *Testing Problems; Test Interpretation; Test Reliability; Test Results; Test Selection; Test Validity

ABSTRACT

The concern with the negative impact on blacks of the national trend toward assessing aptitude achievement and intelligence almost solely through test instruction whose validity and/or administration method is often suspect is reflected in this document. It reports the conference proceedings of the National Association for the Advancement of Colored People. Designed to explore certain issues regarding how testing impacts on public policy, this conference was, however, not designed to resolve all the issues in testing. Specific objectives include the following: (1) to elicit a set of recommendations that seek to deal with the issues and problems identified; (2) to elicit a rationale for the Task Force's recommendations that give meaning to and a basis for interpreting the recommendations; (3) to elicit suggestions for ways to implement the recommendations; (4) to present a summary of all issues considered with their pros and cons. Among the issues discussed are the following: the use and misuse of tests, the psychometric integrity of tests, public policy, and a fair testing code. The common threads running through the report show an awareness that some type of assessment is needed, that tests vastly influence the economic potential of human beings, and that test developers have a responsibility to tell what tests do and do not measure.
 (Author/AM)

 * Documents acquired by ERIC include many informal unpublished *
 * materials not available from other sources. ERIC makes every effort *
 * to obtain the best copy available. Nevertheless, items of marginal *
 * reproducibility are often encountered and this affects the quality *
 * of the microfiche and hardcopy reproductions ERIC makes available *
 * via the ERIC Document Reproduction Service (EDRS). EDRS is not *
 * responsible for the quality of the original document. Reproductions *
 * supplied by EDRS are the best that can be made from the original. *

ED128535

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

NAACP REPORT ON MINORITY TESTING

UD 016361

2

NAACP Special Contribution Fund

MAY, 1976

CONTENTS

EDITOR'S FOREWORD.....	i
PARTICIPANTS IN THE CONFERENCE.....	iv
DISCLAIMERS	v
CONVENER'S INTRODUCTION	1
SUMMARY OF RECOMMENDATIONS	5
REPORTS OF THE TASK FORCES:	
On the Use and Misuse of Tests	7
On the Psychometric Integrity of Tests	13
On Public Policy	21
On a Fair Testing Code	25
ACTION AGENDA	30
SELECTED REFERENCES	32

FOREWORD

Some people seem to think that "intelligence tests" measure intelligence. I've got news for them. "Intelligence tests" do not measure intelligence: they measure one's ability to react to intelligence tests.

I know a man who was tested to determine his dexterity and quickness in seeing relationships between shapes and forms--a very important part of intelligence. On that test, he came out a little lower than a low-grade moron. Then he took a second test, to measure another important component of intelligence, accuracy of memory and quickness of recall. This time, he came out a little higher than a high-grade genius. Was he a moron or a genius?

I once watched two persons taking the same test for accuracy of memory and quickness of recall. They showed startling differences: one ranked very high, the other quite low. The reason was not hard to find: they came from opposite sides of the tracks. The test items were things which had been part of the intimate daily life of one, while the other had never seen and seldom heard of the items mentioned in the test.

Another case in point, directly relevant to the question of alleged "racial" difference in intelligence, was the administration of a test to children in Central Harlem a few summers ago. The children were told to indicate whether a lark was a dog, an automobile, a bird, or a kind of cheese ("Choose one.") None of them answered the question. Mindful of the "LARK" plastered throughout the neighborhood on giant-sized billboards, they later explained their failure on the test question by protesting, "It didn't say 'cigarette'!"

If the matter stopped there, it would be serious enough; but it does not stop here. Not only are individual children tested in order to be detested, standardized tests of intelligence are being used throughout the nation, in ways which permanently affect the self-image of the tested, thereby profoundly altering their own expectations as to their probable futures--and equally affecting the expectations of their teachers. Whole schools, entire school systems, states and regions, are compared to each other on the basis of uniform standardized tests. Racial and ethnic groups are stigmatized. The alleged differences thus "scientifically" validated become self-fulfilling prophecies.

But that's not all. The general mind-set of the nation welcomes these results, because, by training and experience, most people have a need to have somebody else they can feel "better than." We are a highly competitive collection of peoples, living in a nation where competitive sports outdraw all other television audiences. The desire to win is pandemic. From earliest childhood, we are praised for winning--sometimes consoled and often reprimanded for losing. The almost psychotic fervor of a Little Leaguer's parents dominates our culture. And since the schools are part of that culture, each individual is encouraged to out-do his peers. Graded on a bell curve, tested and compared to averages and norms, a child is rewarded or punished by the educational system on the basis of his success in outdistancing others.

"So what's wrong with that!" the true-believing American demands. Much is wrong with it. Of course there is nothing wrong--indeed much is right--in stressing "excellence." But we don't stress "excellence;" we stress "excelling." To be excellent is to shine: to excel is to out-shine. In our national effort to be Number One, we miss the goal of excellence. We settle for its cheap and deceitful substitute, excelling.

That is why a Parent-Teacher Association the other night complained bitterly to a bewildered principal that in every grade half of the children were below the median of their grade. Unfortunately, the general understanding of the meaning of standardized tests is about on the level of that PTA.

The necessity, compulsion, drive to excel others vitiates the more admirable principle of becoming excellent. And because the general public has much at stake in this matter of serving the purposes of a put-down society, the materials in this little Report are important. One doubts that, as a nation, we will quickly put aside our boastful competitive adversary urge to excel; but we ought, at the very least, to observe the elementary dictates of good sportsmanship and fair play. We ought no longer to use standardized tests to announce to the world that some children are "better" than others--particularly when elements of unfairness are built into the testing programs. (Remember the "lark" question for city children.) But even if the tests could be made "culture-free," their use would still be questionable because they are used to predict a child's future, condemning some to be "slow learners"--for life.

This nation ought no longer to permit pseudo-scientists to tell us that some racial groups are inferior to others. We ought not to stand quietly by, permitting the results of nation-wide testing to be used to destroy the self-image of some while falsely inflating the self-image of others.

The men and women who have invested their professional careers in the testing industry are not knaves or fools; but neither are they saints. We should not despair of them. Like all the rest of us, they should be presumed to have the ability to learn. Among other things, this means that we must presume that, given sufficient stimulus, they will put off their old ways and begin to direct their industry toward the production of a new and acceptable product; a testing system designed not to measure degrees of excelling in a competitive rat race but to promote self-understanding in all, as each becomes more excellent. We have no business permitting schools to be run as though life were a rat race. We should be interested in but one race, the human race.

With something of this purpose in mind, the Conference on Testing was convened. The membership of the conference was representative of many points of view. Noticeably lacking, however, was representation from the elitist establishment whose anti-democratic bias makes them perfect practitioners of the art of the put-down. Another unfortunate hiatus in the conference membership was occasioned by the absence of representatives from the two largest customers of the testing industry, the Federal and State Educational

Establishments. Never mind. These others have been having their say and they do not lack a forum. It is our belief that the time has come for another voice to be heard, a voice which speaks without equivocation in behalf of equity and decency and fair play and mutual respect: a voice which will not cease as long as inequity, indecency, unfairness and mutual contempt are thrust upon the lives of our children by elders who should know better.

May I thank the participants in the Conference, even as I apologize for what I have done to their carefully crafted reports from the four Task Forces. In editing, I have endeavored to retain as much as I could of the technical professional language which is natural to them as they communicate with each other; but I have also been mindful of the needs of the general reader. I may have sacrificed something of technical accuracy in order to achieve lucidity. If experts in the field of testing find the language sometimes less than completely professional, they must put the blame not on a supposed incompetence of the Conference participants but upon the tender concerns of an editor who wanted common people like himself to understand what was being said.

BUELL G. GALLAGHER

Vice Chairman, Emeritus,
National Board of Directors,
National Association for the
Advancement of Colored People

May 17, 1976

PARTICIPANTS IN NAACP INVITATIONAL CONFERENCE ON
MINORITY TESTING

- Dr. William F. Brazziel, Professor of Higher Education, University of Connecticut
Mr. Silas Craft, NAACP National Board of Directors
***Dr. Anderson J. Franklin, Associate Professor, The City College of the City University of New York and the Rockefeller University
Dr. William Hayes, Chairman, Association of Black Psychologists
*Dr. Asa Hilliard, Dean, School of Education, San Francisco State University
Dr. George D. Jackson, Professor School of Social Work, Howard University
**Dr. Walter J. Leonard, Assistant to the President, Harvard University
Mrs. Thelma C. Lennon, Director, Division of Pupil Personnel Services, Department of Public Instruction, State of North Carolina
Dr. Arthur L. Mathis, Director, Research and Evaluation, Community Mental Health Center, Menarry Medical College
Dr. Ann Borders-Patterson, University of North Carolina at Chapel Hill
Mr. Silas Purnell, Ada S. McKinley Community Center, Chicago
***Dr. S. O. Roberts, Chairman, Department of Psychology, Fisk University
****Dr. Melvin P. Sikes, Professor of Educational Psychology, University of Texas at Austin
Dr. Madelon D. Stent, Professor of Education, The City College of the City University of New York

RESOURCES CONSULTANTS

- Mr. Lawrence Barclay, Minority Affairs Officer, College Entrance Examination Board
Dr. Edmund Gordon, Director, Institute for Urban and Minority Education, Educational Testing Service
George H. Hanford, Senior Vice President, College Entrance Examination Board
Michael Meyers, Assistant Director, NAACP
Dr. E. Belvin Williams, Vice President, College Board Programs, Educational Testing Service
Julius E. Williams, Director, Armed Services & Veterans Affairs, NAACP
*****Althea T. L. Simmons, Director for Education Programs, NAACP
Dr. Stephen J. Wright, Vice President College Entrance Examination Board

RECORDERS

- Ms. Ana Aponte, Secretary, NAACP
Benjamin J. Colbert, Assistant Program Director, Educational Testing Service
William U. Harris, Area Director, Admissions and Guidance, Educational Testing Service
Ms. Yvonne Ladson, Assistant to the Director for Housing Programs, NAACP
Ms. Esther Spann, Secretary, NAACP

- *Chairman, Task Force on Use and Misuse of Tests
**Chairman, Task Force on Code on Testing
***Chairman, Task Force on Psychometric Integrity of Tests
***Co-Chairman, Task Force on Psychometric Integrity of Tests
****Chairman, Task Force on Social Policy
*****Conference Convener

DISCLAIMER BY THE COLLEGE ENTRANCE EXAMINATION BOARD

The College Entrance Examination Board has participated actively in, and provided partial financial support for, the deliberations which have led to the publication of this report. College Board participation has been with the firm belief that vigorous and continuing efforts should be made to improve standardized tests, to see that they are sensitively administered and accurately interpreted, and to eliminate testing abuses of all kinds. There are, however, several recommendations in the report, including the one pertaining to a moratorium which the College Board cannot support.

Nevertheless, the College Board will give the most serious attention to those recommendations having to do with the improvement of tests, their administration and their interpretation. Moreover, it will continue to be particularly vigilant with respect to any and all abuses.

DISCLAIMER BY THE EDUCATIONAL TESTING SERVICE

The Educational Testing Service has participated actively in, and provided partial financial support for, the deliberations which have led to the publication of this report. Educational Testing Service has participated in the firm belief that vigorous and continuing efforts should be made to improve standardized tests, to see that they are sensitively administered and accurately interpreted and to eliminate testing abuses of all kinds. There are, however, several recommendations in the report, including the one pertaining to a moratorium, which it cannot support.

On the other hand, the Educational Testing Service will give the most serious attention to those recommendations having to do with the improvement of tests, their administration and their interpretation. Moreover, it will be particularly vigilant with respect to any and all abuses.

CONVENER'S INTRODUCTION

The Conference on Minority Testing held in September, 1975 was convened by the National Association for the Advancement of Colored People. It reflects the NAACP's concern with the negative impact on Blacks of the national trend toward assessing aptitude, achievement and intelligence almost solely through test instruments, the validity and/or administering of which is often suspect. This trend which began with the mass objective testing during World War I, has accelerated to the point where in the last two decades great emphasis has been placed on group assessment, which often appears to operate to the detriment of many Blacks.

The assessing of achievement and intelligence of Blacks has been a concern to the NAACP since its founding, for the theme of the NAACP Founding Conference in 1909, was a refutation of arguments that the Negro was physically and mentally inferior. At that historic conference, leading scientists presented papers to refute that widely-held opinion.

During the ensuing 66 years, the issue has been periodically debated, either formally or informally, by NAACP and others, from time to time. The writings of the Moynihans, Jensens and Schockleys and the key Supreme Court decisions in Hobson v. Hansen, Griggs v. Duke Power Co. and DeFunis v. Odegaard have stirred much controversy. More and more, tests have been utilized as the sole, or principal, means of making assessments under circumstances which may adversely influence one's opportunities and achievements throughout one's entire life.

The NAACP has been particularly concerned with the consequences of ability grouping practiced by whole educational systems which results in racial isolation, the enforcement of stereotypes, the labeling of children, and in the reinforcements of feelings of inferiority which can lead to a third class education.

The following vignettes of experiences related to our more than 2,000 units across the country illustrate the scope of the problem:

- Examination scores are used to determine who has access to so-called "examination schools" and who is admitted to the more prestigious colleges and universities.
- Classification systems based on standardized tests label a disproportionately large number of minority children as subnormal and a disproportionately small number of black and other minority children as gifted.
- Many black teachers are kept out of the classroom based on their scores on the National Teachers Examination.
- Many minority group students are unable to enter graduate and/or professional school because of test scores.
- Students are placed in EMR or other special education classes on the basis of test scores.

In 1972 the NAACP adopted a convention resolution calling for a moratorium on standardized tests after other suitable and non-biased criteria for measuring pupil progress and teaching accountability has been devised.

In 1974 the following resolution on testing was adopted by the NAACP convention.

WHEREAS, a disproportionately large number of black students are being misplaced in special education classes and denied admissions to higher educational opportunities,

WHEREAS, standardized tests, e.g., Stanford-Binet and the Wechsler Scale for Children exclude blacks, Puerto Ricans and Mexican-Americans from the representative sample, and,

WHEREAS, such tests label black children as uneducable, assign them to lower educational tracks than whites; deny black children higher education opportunities; perpetuate inferior education; place black children in special classes and destroy growth and development of black children, and,

WHEREAS, students who fail to show a high verbal or numerical ability, score low on the Scholastic Achievement Test (SAT), the Law School Admissions Test (LSAT), the Graduate Record Examination (GRE), etc., and are routinely excluded from college and graduate or professional education,

BE IT RESOLVED, that the NAACP demand a moratorium on standardized testing wherever such tests have not been corrected for cultural bias and direct its units to use all administrative and legal remedies to prevent the violation of students' constitutional rights through the misuse of tests, and,

BE IT FURTHER RESOLVED, that the NAACP calls upon the Association of Black Psychologists to assert leadership in aiding the College Entrance Examination Board and Educational Testing Service to develop standardized tests which have been corrected for cultural bias and which fairly measure the amount of knowledge retained by students regardless of his or her individual background,

BE IT FINALLY RESOLVED, that the NAACP directs its units to use all administrative remedies in the event of violation of students constitutional rights though the misuse of tests and directs National Office staff to use its influence to bring the CEEB, ETS and ABP together to revise such tests.

Following the adoption of this resolution, the Education Department of the NAACP extended invitations to the Association for Black Psychologists, The College Entrance Examination Board and the Educational Testing Service to meet with us to discuss the concerns of our Association regarding testing. During the ensuing year, representatives from these organizations met with us on three occasions to identify and focus on selected key issues. The results of our deliberations led us to convene the Conference on Minority Testing.

This Conference was not designed to resolve all the issues in testing. Rather it was designed to explore certain issues regarding how testing impacts public policy; whether or not there should be a code on testing and if so, what should the code encompass; the use and misuse of tests and the psychometric integrity of tests. The specific objectives as outlined to the conferees, was to elicit:

1. a set of specific recommendations that seek to deal with the issues and problems that have been identified.
2. a rationale for the Task Force's recommendations that give meaning

to and a basis for interpreting the recommendations.

3. suggestions for ways and means to implement the recommendations.
4. a summary of all issues considered with pros and cons.

The relevance of the issues addressed during the Minority Testing Conference was underscored in many ways. To mention two, we can point to the fact that participants at the Conference were educators, representatives from the testing industry, professional and community organizations and, secondly, we noted the many common threads running through the four Task Force reports.

Those threads include an awareness that:

- Some type of assessment is needed.
- Tests, in our credential-oriented society, vastly influence the economic potential of human beings.
- There is a need for widespread dissemination of information regarding testing in a form easily understood by all segments of the population.
- Test developers have an advocacy role to perform including sanctions for continued abuses.
- Test developers have a responsibility to tell what tests do and do not measure.
- Persons who take standardized tests must know their rights.
- Guidelines for the administration of tests should include specifics regarding the type of environment necessary for optimum test performance.
- Subjects and users must understand what is expected.
- Criterion-referenced approaches and materials should be further investigated.

In view of the fact that tests, (since they stratify or certify people) are, in effect, married to national policy issues, and determine what kinds of people, from what background, where they will fit in the society, what role they will play, and the like, it is incumbent on the Minority Community to galvanize educators, the testing industry, parents, students and community organizations to work systematically to insure that the assessment of individuals is culturally fair.

This is vital if all Americans are to enjoy equal opportunity in every aspect of public life.

The NAACP expresses its sincere appreciation to the Task Force Chairmen and the support staff.

We are also grateful to the individuals and organizations that contributed materials and data as background information and to the College Entrance Examination Board and the Educational Testing Service for the financial assistance which made this Conference possible.

ALTHEA T. L. SIMMONS

Director for Education Programs
The National Association for the
Advancement of Colored People.

May, 1976

SUMMARY OF RECOMMENDATIONS

1. That there be a moratorium on all current standardized tests unless such instruments conform to recommendations set forth in this Report.
2. That a national monitoring body be established with the power to enforce, through sanctions, to assure proper assessment and policy regarding the administration of assessment tools.
3. That companies that develop, publish and sell tests assume (or continue to assume) major responsibility for assuring the correction of the deficiencies in their instruments.
4. That the testing industry be held responsible for the development of assessment procedures which conform to professional standards as described in Standards for Educational and Psychological Tests developed by the Joint Committee of the American Psychological Association, American Educational Research Association and the National Council on Measurement in Education.
5. That the testing industry, at a minimum, include within the information it publishes concerning standardized tests of ability, achievement, personality and any other assessment procedure, specific data regarding predictive, content and prescriptive validity.
6. That, where standardized assessment results in the disproportionate sorting of groups according to ethnicity, the test developer provide separate validity coefficients for ethnic groups to which the assessment procedure is to be applied.
7. That test developers describe the probable main effect-variables in the instructional setting, in standardized terms, which must be considered along with the results from standardized testing if the interpretations of results are to be meaningful or acceptable. For example, the ethnic background of the examiner is known, in some cases, to affect the scores which children earn on IQ tests; therefore information about the ethnic and other important background factors of the examiner should be reported simultaneously with the test scores.
8. That the testing industry establish and fund an independent research and development corporation charged with the responsibility (1) to identify the critical problems in assessment as they relate to minority groups; (2) to sponsor research to investigate those problems requiring study; (3) to sponsor appropriate development work; and (4) to involve researchers who have the endorsement of minority professional and community associations, and that a minimum sum of four per cent (4%) of the income over expenses for non-profit testing corporations and the sum of four per cent (4%) of profits for profit-making testing corporations be set aside in support of the above objectives.
9. That test publishers exercise an advocacy responsibility which require that test objectives be stated clearly, that the process be fully described so that subjects and users understand exactly what is expected and how it will happen. This is the principle of informed use and informed consent.

10. That publishers of tests state, with clarity, on all descriptive information concerning a test, the specific uses for which the test is designed, the specific limitations of the instrument, and that they provide information as to how the results should be interpreted in acceptable professional practice.

11. That we call upon the appropriate professional associations, specifically the Association of Black Psychologists (ABPsi), American Psychological Association (APA), the American Education Research Association (AERA), the Association for Non-White Concerns of the American Personnel and Guidance Association (ANWC), and other appropriate groups to establish minimum standards for those who administer and interpret such standardized tests as tests of intelligence, aptitude, achievement and personality, and to develop standardized basic information about test administrators and the testing environment, to be provided routinely with any test scores.

12. That the NAACP mount a concerted effort to identify instances of testing abuse which call for legal remedies.

13. That the NAACP establish a national task force for the purpose of developing specific guidelines for laymen's participation in and support of any standardized assessment procedures, and that the NAACP insure dissemination of the guidelines to the broadest possible audience.

14. That citizens urge state elected officials to pass legislation to establish a task force for the development of an independent Office of Consumer Affairs for Testing and Student Evaluation.

15. That the Association of Black Psychologists design and conduct workshops around the Task Force Reports which will include, but not be limited to, the effects of race of examiner on test taker; differential validation and reliability; improper use of IQ as dependent variables in research projects, biases in test construction; the problem of misinterpretation and the development of alternative means of assessment.

ON THE USE AND MISUSE OF TESTS

Competent scholars have always recognized the possibility that intelligence tests and aptitude tests might be abused. Beginning with Binet, and including such later scholars as Ellis, Klineberg, Allison Davis and such contemporary scholars as Robert Williams and Leon Kamin, they have called attention to potential and actual abuses. Nevertheless, the abuses have continued. Many black and other minority children have been unfairly stigmatized or inappropriately grouped for instructional purposes. Culturally sophisticated teachers and psychologists have protested that children who have been thus identified as "less able," "retarded," or what might be called "the six-hour mentally retarded," seem to lose that retardation immediately at the end of the school day. But back in the classroom the next morning, they conform to the predictions of those who first sequestered them in "special" classes.

The racist history of the testing movement is documented in Leon Kamin's The Politics and Science of I.Q., as well as in the writings of others. (See List of References appended). The current misassessment of blacks and others is rooted in a long--and often unscientific and malevolent--history. In this Task Force report, we are concerned with tests of "aptitude," "achievement," "personality,"--both individual and group--and any other method of assessment of individuals which results in ranking, sorting out and inviscerally comparing.

It should be the function of assessment, including testing, to facilitate the development of human resources. To be acceptable in the field of education, testing programs should reveal student growth in skills, attitudes and understanding. To be used in connection with selection for employment, tests must measure variables which have something to do with the job for which an individual is being tested. These are the criteria against which all tests and testing programs must initially be measured: every professional testing procedure, method, technique, instrument or material--and the environment in which the tests are administered--must conform to these standards. That is the initial judgment by which any testing program must be evaluated, even before it is put into use.

It is important to note here that we are very much in favor of, and recognize the need for, appropriate and competent assessment. However, the existing standardized tests are unacceptable. They fall short of meeting three fundamental criteria (in addition to that of relevance, noted in the previous paragraph). Tests and testing procedures must predict accurately what they promise (predictive validity). Tests must measure adequately the content of the area they purport to cover (content validity). The testing program must be capable of leading to prescriptions which result in positive growth for the person being tested (prescriptive validity). It is abusive to misassess with an inadequate instrument. Equally important, it is abusive to continue, year after year, to use testing programs which have proved themselves to be tools which are either irrelevant to student progress or which

actually, by predicting failure, induce malevolent results. Current aptitude testing or other testing processes and practices which result in the misassessment of blacks and other minorities result in educational mistreatment. All such tests must therefore be examined, not only for cultural or ethnic bias, but more importantly, in the light of their intended ultimate use. For all these reasons, we assert that fundamental questions must be asked about testing, beginning with the motivation of the movement and continuing through the assumptions which underlie the use of the instruments, raising questions about the accountability of the testing industry when measured against the educational outcomes for students.

In the light of the foregoing general considerations, the following recommendations are made:

RECOMMENDATIONS TO THE TESTING INDUSTRY

We recognize that a number of theoretical problems in assessment need attention. Among these are the issues of norm versus criterion-referenced testing; predictive versus prescriptive validity; aptitude versus achievement testing; content versus contextual validity; status versus process assessment; and the extent to which some less readily measured goals will be disregarded if standardized testing is given principal priority and support. (Knowledgeable members of the teaching profession will recognize that each item in this list calls for book-length treatment).

In addition, underlying all of these theoretical issues is a basic one, the question of "universality" in human behaviour. When a single standardized test is administered to an entire population, the underlying assumption must be that it is fair to all because all are "alike" in their possession of the qualities and information to be tested: they differ only in the degrees to which they possess these qualities or have the information. Such an assumption cannot legitimately be made for the entire population of the United States. In the light of the information that we have about differences between and within cultural groups within this nation, we believe the assumption of "universality" to be a serious error. Since there is no standard experience, the "standardized" test poses a serious problem of theoretical difficulty. When applied to the problem of testing minorities, this difficulty is greatly increased.

The foregoing concerns by no means exhaust the list; but this weekend conference can do no more than suggest the magnitude of the problems before us, without presuming to include everything. We know that existing tests have been built at costs that run into millions of dollars; yet, even with such an investment, we do not have tests that work as they are supposed to work. The truth is that the task of devising a full remedy is beyond the competence of a weekend conference, but the burden of proof for the utility and satisfactoriness of any testing procedure rests with the producers. We will have discharged our function in this Task Force if we point to inadequacies and indicate general directions of needed improvement.

The testing industry is (or ought to be held) responsible for the development of assessment procedures which conform to professional standards. These standards are provision-

ally described in Standards for Educational and Psychological Tests published by the American Psychological Association Inc. in 1974.

Within these general prescriptions, the minimum responsibility of the testing industry is to include within any standardized tests of ability, achievement, personality, or any other assessment procedure, specific consideration of and data related to three kinds of validity--predictive, content, and prescriptive (as we noted above). These data must be expressed as a coefficient or other appropriate systematic expression which is developed as a result of an adequate validation study. Such data must accompany all administrative manuals for use with standardized tests. Furthermore, in instances where standardized assessment results in a disproportionate sorting of groups according to ethnicity, the test developer must provide separate validity coefficients for the ethnic groups to which the assessment procedure is to be applied.

In addition to the foregoing, it is also the responsibility of the producers of assessment procedures to describe probable main effect-variables in the instructional setting, in standardized terms, which must be considered along with the results from standardized testing if the interpretations of test results are to be meaningful or acceptable. For example, we know that the ethnic background or level of skill of those who administer or interpret tests often has a major effect on the announced results of the program. Therefore, we have the right to expect that the testing industry will devise satisfactory and systematic ways of taking such effects into account.

Agencies make money from the administration of tests. Therefore, they have the responsibility to finance the measures necessary to the correction of their testing programs. Among these measures must be the inclusion, within the test-construction process itself, of persons drawn from culturally diverse backgrounds and of various ethnic identities. The evaluating experts must include representatives who are acceptable to the minority professional groups and other major community groups if the testing program is to escape its present image as being unduly weighted in favor of the dominant forces of American society.

The testing industry must establish and fund an independent research and development corporation charged with the responsibility (1) to identify the critical problems in assessment as they relate to minority groups; (2) to sponsor research which will investigate these problems; (3) to sponsor appropriate development work; and (4) to involve researchers who have the endorsement of minority-rooted professional and community associations.

In view of the enormous profits which have been made over the years by the testing industry, we recommend that some small redress of past errors be made by the voluntary application of four per cent (4%) of the net income of non-profit testing corporations and four per cent (4%) of the net profits of profit-making testing corporations, to the support of the foregoing objectives.

Our next point is that the makers of tests must be accountable for the uses to which their tests are put. Test publishers must be responsible for monitoring the use of their

tests by assuming an advocacy role when necessary. In instances where tests are used (or are about to be used) without due observance of this advocacy role, test publishers should apply sanctions, including the denial of the use of their product by those who misuse it.

Moreover, test publishers have an advocacy responsibility which requires that test objectives be clearly stated, that the process of administering the test be fully described so that subjects and users both will understand exactly what is expected to happen and how it will happen, in non-threatening, affirmative terms. This we would call the principle of informed use and informed consent.

Finally, publishers of tests must state with clarity, in all descriptive information concerning a test they publish, the specific uses for which the test is designed, the specific limitations of the instrument, and full explanation as to how the results should be interpreted.

All of the foregoing steps must be taken by the testing industry itself. The broadest interpretation possible should be made of the concept of abuse, within the field of standardized testing, since abuse can occur at any and all points of the testing process, from initial development and conceptualization through utilization and interpretation.

We assert here that the misuse of tests, whether due to ignorance or to bias or to indifference, is an important factor in the total problem we are discussing; but we also assert that it is by no means the whole problem. Inherent in the instruments themselves, particularly in the instruments designed to measure aptitude, is a basic bias which must be corrected at its source. That source is within the offices and work-rooms of the testing industry. No amount of training or orientation of users of the tests will correct this built-in bias; but a successful effort to correct such bias will result in high predictive-, content-, and prescriptive-validity.

RECOMMENDATIONS TO PRACTITIONERS

Valid standardized tests, even when they are produced by the industry, can still be used in error, or their full potential lost. Appropriate use would include diagnostic and prescriptive procedures which lead to pupil gains. In the instance of educational tests designed to measure achievement, recent innovations in standardized assessment seem to offer possibilities which may be supportable.

Some recently devised tests would seem to have a greater utility for use in the school room than earlier tests. Those with the greatest utility have a closer relationship to the instructional process, offering valuable insights as to teaching strategies. These tests are often referred to as "formative-summative" tests. A common use of these tests occurs in reading and mathematics programs, where short "locator tests" are used to ascertain proper beginning points for individual students in appropriate units of instruction, and "summative" tests are used to ascertain the degree of mastery of materials and the readiness of the child to move on a more advanced level.

While no specific existing tests can be fully endorsed at this time, the principle involved in the assessment procedure just described appears to merit our support.

Practitioners are reminded that abuses in testing programs often occur with reference to majority as well as minority children. While we stress the pattern of abuse with reference to minority children, because the magnitude of the error is greater, we do not overlook the welfare and progress of the majority child--who should also benefit from a testing program designed to meet the criteria and serve the purposes we are discussing.

Any test which results in a disproportionate distribution of students on an ethnic basis must demonstrate that this disproportion does not come as a built-in error due to low predictive-, content-, or prescriptive-validity. Where such a conclusion cannot be established, persons asked to participate in the use of the test should decline to do so.

Perhaps an example will help to explain what we mean. Persons familiar with the processes of test construction will admit that items which, in the preliminary testing and validation, appear to differentiate between males and females are thrown out, presumably on the assumption that there are no real differences in intellectual functioning as between the two sexes. Yet items which appear to differentiate between whites and blacks, in the same battery of tests, are retained--presumably on the assumption that there are real differences between the races. Thus, the test which is published and administered, after the preliminary run, carries no built-in bias as to sex but does carry a built-in bias as to race. The fault is clear. The correction ought to be equally clear.

We call upon the appropriate professional associations, specifically the Association of Black Psychologists (ABPsi), the American Psychological Association (APA), the American Educational Research Association (AERA), the Association for Non-White Concerns (ANWC), of the American Personnel and Guidance Association, and other appropriate groups to establish minimum standards for those who administer and interpret standardized tests such as tests of intelligence, aptitude, achievement and personality. These standards must be equivalent in rigor to the standards which have been established for publishers of standardized tests and diagnostic techniques. Further, we call upon these same associations to develop standardized basic information about test administrators and the test environment, to be provided routinely with any test scores.

RECOMMENDATIONS TO THE NAACP

Recent court decisions (Larry P. et.al. v Wilson Riles, et.al.; Diana, et.al. v California State Board of Education) have been won on the principle that predictive validity for a standardized test was lacking in the instances invoked. In the case of Griggs v. Duke Power Co., the case was won on the principle that employment tests must be "job related". To maintain the momentum generated in these cases, it is recommended that other instances which may appropriately be pursued to a legal remedy be immediately identified and pursued.

Particular attention should be paid to the results flowing from the use of the National Teachers Examination. A disproportionately high number of those failing this examination

come from the minorities. Teaching certificates are denied or credit for promotion is withheld as a result. Yet there is little, if any, evidence to indicate that these examinations have any predictive relationship to the jobs sought. The principle of Griggs v. Duke Power Co. would appear to apply.

If predictive validity cannot be demonstrated by the Law School Admissions Test (LSAT), the Graduate Record Examination (GRE), the Medical College Admissions Test (MCAT), the Graduate Management Admissions Test (GMAT), and other screening devices used in connection with admission to graduate and professional schools, technical schools or other institutions of higher education, legal redress should be sought.

We call upon the NAACP to establish a Task Force for the purpose of developing specific guidelines for the participation by laymen in the effort to devise satisfactory assessment procedures. Further, we call upon the NAACP to insure dissemination of acceptable guidelines to the broadest possible audience.

RECOMMENDATIONS TO LAYMEN

No person should consent to participate in an assessment program unless he or she has a reasonable understanding of the procedures to be applied, and of the outcomes to be expected.

RECOMMENDATIONS TO OFFICE HOLDERS

We urge that the public obligation to protect consumers against improper exploitation and victimization be recognized by legislative action to establish, outside the educational bureaucracy, a properly staffed Office for Consumer Affairs (Testing and Student Evaluation).

ON THE PSYCHOMETRIC INTEGRITY OF TESTS

INTRODUCTION

The purpose of the Task Force on Psychometric Integrity of Tests was to highlight a number of questions and issues central to the technical development of tests. There are many techniques and procedures in test construction that must be followed to ensure that an appropriate measure has been devised. Although most of us know very little of the technical skills required in building a house, we place faith in the constructor's ability to provide a safe and secure home. If too many flaws occur in its construction our lives may be placed in jeopardy and we must question the ability, perhaps the integrity, of the constructors. Over the years a plethora of regulations have helped govern the construction industry as a guaranteed protection of the citizen from fraud and misrepresentation of quality of product. Comparable to the building-trade industry is the test-construction industry, whose technical procedures must also be scrutinized for quality of product. Decisions made using test results oftentimes place the lives and futures of children and adults in jeopardy. That is to say that test data must provide constructive information which will enhance the personal development of all individuals regardless of group membership. To date tests have principally been used as instruments for screening individuals and placing them in classes of opportunities. However, tests must also safeguard the individuals' potential and capacity to grow, if not specifically reveal ways in which they may be developed (e.g., in perhaps the diagnostic-prescriptive manner). This Task Force therefore raises questions about important technical areas of test construction which must be considered in the totality of issues on cultural fairness in testing for minorities. These questions represent areas for more complete analysis and discussion.

The Charge to the Task Force listed a number of questions which must be fully addressed by competent experts in the technical procedures of test construction and utilization. Our response cannot be definitive, given their complexity and our shortness of time. However, our intent is to share with others our awareness that part of the problem in fair testing of minorities rests with the technical development of tests as well as their use. Each issue, therefore, will be treated separately with a statement reflecting, in our judgment, an expansion of the issue in terms of minority concerns as well as suggesting the areas which need further inquiry.

1. As in most things, tests should have a reason and purpose underlying their development. Although attention may be given to the purpose for which tests are used, there is an equally important problem to consider--namely, the theoretical assumptions about human performance (usually intellectual performance) and about how the test format purports to measure and actually does measure the areas of interest. For example, one would assume that to test for intelligence a concept of what "intelligence" is and its behavioral manifestations would be clear. Technically, it is not. Intelligence has become synonymous with scores on tests of intelligence (i.e., the Intelligence Quotient or IQ). The full range of

intellectual potential and capacity of humans is still unknown. We know even less about how humans process information. Therefore, it is a gross injustice to discredit a person's present intellectual capacity, much less potential, solely on the basis of test scores which are in turn frequently based on a fragile foundation of theoretical understanding and empirical evidence about human intellectual ability. The tendency is to take what the majority of people can do as indicative of ability levels. However, the presupposition is that all persons have had at least equal access to the experience central to the skills of the majority. Nevertheless, within the issue of theoretical concepts explaining behavior, we must ask whether these basic assumptions consider the breadth of life contingencies (particularly minority human development within a racist society) which might affect intellectual growth. Certainly minority scholars have had little opportunity to participate in this domain of scholarly speculation, much less positive reception and support of their ideas and assumptions about human performance.

It is therefore necessary to scrutinize fully the assumption underlying many of these tests to see whether the root of unfairness in testing stems from the initial conceptualization of the behavior one is trying to measure, and ultimately how it is incorporated into a standard test format.

2. Many tests, particularly intelligence tests, have several parts (sub-tests) which, when taken together, are supposed to represent a total picture of a person's abilities. The same stringency of controls employed, in general, for the overall development of a test applies also to the establishment of criteria in the selection and use of parts of a test. For example, commonly used subtests in a test require demonstration of one's verbal, quantitative, or performance abilities. Certainly we agree that these components of ability are important in intellectual competence, but we are also concerned about the tendency for test developers and clients to neglect other domains of human capacity which affect performance (e.g., motivation, personality, emotions, memory strategies, etc.) Our concern, therefore, is to question the exclusivity of priority areas in ability testing and to emphasize the need to broaden the areas of human capacities considered.

3. Another area of important concern to us is the process of determining test questions which are ultimately selected to be a part of the instrument. It is acknowledged that this is an arduous process of sifting through numerous questions until a final group of items best represent the purpose of the test. Moreover, it is acknowledged that the statistical process involved has immense utility. However, the judgments in the development of questions and the interpretation of statistical data in the process of selecting items is still a human one, it can be no better than the competence and sensitivity of the experts assigned to this stage of test construction. Therefore it is important that at this juncture of test development we are clear about the criteria of the excellence we are trying to measure, and how widely that excellence is represented within the pluralistic cultural milieu of American society.

It is precisely the fact that American society is culturally pluralistic which concerns

us, in the item selection process as well as other stages in test construction. Many items are selected on the basis of their alleged ability to measure the skills of the average person, the population in general. Very often this is at the expense of those items which differentially distinguish between various groups. In other words, a question which blacks may do well on, but not necessarily other groups represented in the tested population, may be excluded from the final form of a test because of its lack of representativeness within the general population. This may also be true of any other group distinctions one wants to make. The fact is that in the pursuit of questions which represent the general population, we may be overlooking information which has a comparable potential for representing human capacities, but in a selective manner. Moreover, excluding items from a test which favor one group but not necessarily another, may be placing in jeopardy the representation of that group's ability and potential.

4. The factor analytic procedure is a statistical method for finding out how a pool of items cluster together or are truly independent. It is a procedure central to the item selection process in test development. The arguments for and against this procedure and its dominance in determining test items is recognized as too technical to entertain within this document. The important point, however, is the recognition of the fact that this procedure has limitations; and, secondly, that there are few minority professionals involved with this procedure of test development as participants in judging the use of this statistical technique in the best interest of minority groups. There is a critical need to determine how much this statistical procedure may influence the type of test items selected, and what impact decisions by statisticians may have on the way minority ability is profiled on standard tests.

In addition, since there is considerable dependency on this procedure in test construction, we again question whether there are not other methods which might be comparable in objective, but more sensitive to the way different groups exhibit their abilities. In essence, we request a broadening of methods of test construction and assessment, particularly adopting those which would provide a fair representation of individual and group skills.

5. One area of testing which has been controversial is the issue of standardization procedures, the development of norms for a given population. Essentially, this involves profiling the range of competence demonstrated by a population on a set of questions comprising a test. These are ultimately refined into standards to which levels of individual achievement are to be compared. Central to this procedure is to profile the range of test performance from a representative assortment of individuals and groups. It has always been a concern of minorities that as a group we have not been adequately represented in the normative populations. Recently, some testing agencies have attempted to revise their norms by including a broader representation of minorities in the sample (e.g., Weschler, Intelligence Scale for Children, Revised (WISC-R).) There is still controversy as to whether proportionate representation will resolve the issue of fairness. The selection process should include those factors (e.g., socio-economic and geographical residency) that are

appropriate for the representative group which is to be tested. If the major concern is to have a test which is representative of the general population, the majority group in this country (whites) will prevail in the performance results of the test. Consequently, the peers to whom one's ability is compared are essentially representative of the standards set by individuals or groups outside the target populations.

6. Two major concerns about tests are whether they are reliable and whether they are valid. By reliability is meant the consistency of a person's performance on a test if repeated; that is, confidence that the level of an individual's or group's performance will not vary significantly the second time around. Validity of a test focuses on the truth or accuracy of test results, the degree of faith that can be placed in the test as measuring what it claims to measure. The importance of these dimensions for the issue of test fairness for minorities rests with the interdependency of these two factors. A test should be both highly valid and highly reliable. This is not always the case. A test with little validity may be reliable (i.e., one's performance may be consistently high or low in a test-retest circumstance), but the content or predictive capability may be unrelated to the test objective or the desired basis for judgments. Parenthetically, this is an issue in the IQ controversy.

Similarly, a test may wrongly be considered valid if it meets only one or some criteria--that is, if it has content validity, predictive validity, concurrent validity and/or construct validity. In the testing controversy, discussion tends to center on predictive and content validity. The issue of "culturally biased tests" principally refers to the fact that the contents of tests are not representative of the socio-educational experiences of minorities. The predictive validity of any test can really be no better than the defined future for which probable success is being predicted (i.e., success in school or on the job). The numerous issues within the meaning and significance of validity and reliability of tests (particularly as germane to the use and abuse of tests) requires that this area not be treated lightly. The public must fully understand the consequences in the assessment capabilities of a test when it does not meet acceptable standards of validity and reliability.

7. Cut-off scores and criterion variables are two other important factors in the testing process. By cut-off scores is meant the point at which a person's performance (a test score) permits a judgment about acceptance or rejection, that is, his or her ability or potential in terms of likelihood of success. Related to cut-off scores is the criterion variable or the performance goal. This refers to skills which a person or institution establishes as representative of acceptable performance. For example, a criterion variable for college admissions is "potential grade point average". This is purported to be indicative of college success. It may be determined that scores at a particular level on a test correspond to a projected level of performance in college (i.e., expected GPA). Consequently, as in the case of admissions tests, e.g., College Boards (SAT) or Graduate Record Examination (GRE), a person's score supposedly serves as a predictor of the level of achievement he or she will attain. By looking at the relationship between the admissions tests scores of students who

are successful in college, schools establish their "cut-off scores" as well as criterion variables.

The Task Force's difficulty with this process is that prediction models of potential success build a dependency on two factors -- the test score and the institution's definition of "success" -- with little consideration for other factors which might contribute to a person's success (e.g., motivation). As previously noted, if there are serious questions about the validity of a test, then the significance of the test scores is brought into question. Furthermore, if test scores are used in a process of selection based on prediction, the criteria of "success" depends upon the institution's valid appraisal and inclusion of all salient factors that represent competent performance. Very often institutions provide a narrow definition of success or skills required (e.g., success in college equals potential grade point average).

Many minorities are denied access to educational opportunity because of institutional dependence on poorly determined levels of cut-off scores, or are enrolled in institutions where, if traditional criteria of selection had been employed, they would have been excluded from this opportunity (a case in point is the City University of New York's Open Admissions Policy).

8. Because standardized tests are reasonably short, economical and easy to administer, they have become the most convenient method of assessment. The Task Force's concern is that this dependency on such a structured format has limited the exploration of other means of assessing intellectual ability. We are concerned not only with the limited domain of demonstrated competence on which tests tend to focus, but also the limited strategies of assessment (e.g., standardized pencil and paper tests) employed by institutions. It seems unfair to judge the obviously broad capacities of human growth and development in such a narrow manner. Likewise, given the breadth and variety of human social experience it seems unfair to build dependency on measures which actually test only for degrees of conformity (in knowledge and experience) among the general population -- not innate abilities. The judgments about one's ability and, subsequently, the consequence for one's life are too important to be limited to performance on tests. Other alternative means for measuring and predicting human potential and capacity must be developed and employed. The multitude of assessment strategies must reflect the pluralistic compositions of American society. This is no easy accomplishment, but at the same time this objective must not be ignored at the expense of someone's life opportunities.

9. The importance of how a test is used is an enormous issue in and of itself. Within this Report, another Task Force is discussing this problem alone. Although the content of that Task Force's Report covers this issue in depth, it cannot be overemphasized that the qualifications of the examiner or test user are equal to, if not more important than, the tests which are used, since even the best instrument in the hands of unqualified users can lead to disastrous results. No mechanism exists at present to insure that only well-qualified persons (or agencies) will use the tests or establish policies and procedures based upon test results.

10. The Task Force on Use and Misuse of Tests speaks specifically to the current objective of "criterion-referenced" tests. Therefore, this Task Force will address itself to a related assessment methodology, i.e., development of tests which are a derivative of the social and cultural experiences of the specific group to be tested. Tests of this nature are considered "culture-specific tests". The utility of these tests has been hotly debated; but in the Task Force's judgment they serve a purpose -- if only in delineating the context and content of learning experiences for that particular group. In this regard, it is believed that efforts to develop "culture-specific tests" as an inductive procedure ultimately advances our knowledge from specific bases of information to those common features which describe the general population. With the current test-industry effort to find the domains of convergence in excellence and performance of a general population (which in effect standardized tests do), little weight is given to the idiosyncratic nature of learning. It is this Task Force's belief that tests standardized both for the general population and for specific populations add to the information on variations in the acquisition of knowledge. Within a pluralistic society, we must know what people have in common, but also where they differ, without discrediting either in the quest for understanding. To seek a test which is "culture-free" in content is nice in theory, but to date has proven impractical. There is very little that can be identified which is not influenced by the cultural context in which it is nurtured and expressed. In general, the Task Force endorses the efforts to develop and refine criterion-referenced and culture-specific tests as part of alternative assessment strategies.

In conclusion, the Task Force on the Psychometric Integrity of Tests recognizes that there are many technical issues involved in the construction of tests. These must be considered within the context of their impact on fairness in testing minorities. Because of the complexity of many of the areas discussed within this report, it should not be construed that all points have been raised -- much less exhausted. The Task Force has tried to highlight what constitutes some of the major concerns presented to minorities by the technical process in test construction. As in so many other professional areas, minority professionals are badly under-represented in the testing industry. In particular, the contribution of minorities within the institutions and agencies which develop tests and set policy has been negligible. The test development process has remained untouched by an external system of public accountability. The consequences to one's life from performance on tests can be as pivotal as any lifesaving drug. In the latter instance, there is public accountability (i.e., FDA), in the former, there is none. The following recommendations of this Task Force are a step in the direction of public accountability of the testing industry.

RECOMMENDATIONS

The Task Force asserts that the basic goals and purposes of assessment are as appropriate for minority individuals as for persons in the majority group. It is the consensus of the Task Force that there can be a meaningful and worthwhile place for the testing function in the assessment of individuals. However, the issues which concern the NAACP, and

in fact, the purposes of this Conference, relate to the constraints and restrictions placed on minorities by testing; by the fact that those negative results preclude access to educational and occupational advancement; by the absence of social and cultural considerations in test construction; and by individual and group values which affect test performance and interpretation.

The Task Force's interpretation of the Resolution adopted by the 65th NAACP Annual Convention, is that it serves as a means for identifying the need to investigate the concerns of minorities in testing. It has also provided opportunity for study which may devise better ways of developing and using test data.

SPECIFIC RECOMMENDATIONS

1. Normative procedures and specification must be carefully developed to ensure fairness to the test-taking population. Information of this type should not be used for minority groups or individuals without appropriate norming (study) on that population.
2. The test development process must consider the different cognitive structures and styles of different groups. Studies related to minority test performance indicate that group differences do affect performance. Factor analytical methods may yield information relevant to understanding these cognitive structures and styles, but there still remains the problem of how these factors are related to effective performance.
3. The test selection-predictive system should include other variables, e.g., motivation, persistence, "creativity", and other personality measures. There is frequent mention of these important factors in the discussion of minority assessment, with little follow-through.
4. The relationship between the time factor and test results should be expanded with minority groups. This exploration should not be limited to test speediness, but also to the length of time covered by the criterion measure.
5. The test administration process, both for individual and for group testing situations needs to be monitored to ensure quality control of tests results. We recommend that guidelines be developed that will provide opportunities for optimum test performance.
6. Culture-specific tests should be considered an integral part of the test construction process.
7. Noting the difficulty of establishing the relevant criteria, particularly where predictive validity strategies are used, it is strongly recommended that culturally-appropriate and content-valid, criterion-

referenced procedures be investigated and adopted. Some guidance should be provided in the development of criterion variables, e.g., tests for firemen, policemen, etc.

8. The development of alternative strategies in assessment is strongly encouraged.
9. There should be a thorough review of current tests being used. NAACP, ABPsi, and others should actively pursue the possibility of seeking funding sources for such a project.
10. There should be a concerted effort by testing industries and professional schools to recruit and train more minority persons in psychometric techniques. Furthermore, testing industries should increase participation of minority professionals in the test development process.

ON PUBLIC POLICY

INTRODUCTION

We live in a highly diverse, highly competitive, credential-oriented society where success and winning have become so prized that often they become ends within themselves. As a result of the frenetic milieu created by this kind of attitude and philosophy on the part of most Americans, an increasing number of assessment tools and techniques are being employed to facilitate the classification, stratification and certification of individuals and groups in our society.

The fact that testing and the results of testing (assessment) have had (and continue to have) a strong impact upon public policy is not new. A review of litigation involving the alleged misuse of tests by certain businesses and certain industries and the results of judgments in those cases indicated clearly the role that testing has played in hiring practices and the impact that they were seen to have on policy. The cases to which we allude here involved white plaintiffs.

The determination of public policy related to the testing of minorities (and more particularly American blacks) has been insidious and extremely deleterious. The history of our country is replete with evidence of so-called scientific material designed to show the inferiority of the Negro. The Police Reference Notebook states, "A large body of literature came into existence to prove that the Negro was imperfectly developed in mind and body, that he belonged to a lower order of man, that slavery was right on ethnic, economic and social grounds..." The rationale for counting slaves as three-fifths of a person for determining the number of representatives that a state might send to Congress was based on "scientific" tests that "proved" the inferiority of the Negro. For centuries blacks were disfranchised by the use of "tests." It is unnecessary to belabor these historic truths; it is important to realize, however, that testing helped weave the racist fabric of the United States. There are daily, painful reminders that white racism (the unfair treatment of non-white persons, based solely on skin color) continues to flourish in the United States. The insidious character of white racism (and the most dangerous component of its insidious nature) is most harmful when it becomes an inextricable part of test construction. This situation can obtain without the conscious participation of the developer or publisher. Nevertheless the disastrous results, however unintentional, are assured. When public policy is based on an already abusive instrument and the interpretation of an insensitive researcher, persons belonging to minority groups are exposed to a variety of inequitable and unethical behaviors and treatments.

THE PROBLEM

The Task Force on Public Policy, in its concern about the impact of tests on public policy, felt it incumbent upon its members to consider various facets of the problem. That

is to say, the person who uses the test and the individual who formulates public policy are as important as are the tests, in the overall consideration. Naturally, the publisher must assume his part of the total responsibility for fair and equitable public policy.

Some of the psychometric instruments having the greatest impact upon public policy have been:

1. Tests of "Intelligence" (I.Q. Tests)
2. Personality Tests
3. Placement Tests
4. Achievement Tests
5. American College Testing Program (ACT)
6. Scholastic Aptitude Test (SAT)
7. Miller's Analogies
8. Graduate Record Examination (GRE)
9. Interest Inventories
10. Graduate Management Admissions Test (GMAT)
11. Law School Admissions Test (LSAT)
12. Medical College Admissions Test (MCAT)
13. Other Admission or Aptitude Tests

Certain questionnaires have also been responsible for questionable public policy, but this Task Force could not consider them as a part of this report.

RECOMMENDATIONS

After careful study of the charge to the members of the Task Force, and following in-depth deliberations, the Task Force on Public Policy makes the following recommendations:

1. There should be a moratorium on all current standardized tests, unless these instruments conform to recommendations submitted by the other Task Forces that comprise the Conference on Minority Testing.

This recommendation is based on studies which tend to prove that most standardized tests are inherently racist (intentionally and unintentionally) and that they do discriminate against minorities. The precedent for calling for this moratorium has been set by (a) the National Association for the Advancement of Colored People (NAACP), (b) the National Education Association (NEA), (c) the Association of Black Psychologists (ABPsi), and (d) the National Association of Elementary School Principals (NAESP).

This call for a moratorium does not mean that all forms of assessment should be abolished. Some form of assessment is and will always be necessary. It does recognize the damage that has been done and is being done as a result of public policy based on the spurious results obtained by the use of questionable instruments, as well as by men whose motives and preparation for their work are questionable.

The over-representation of minorities in Special Education classes across the country and the many cases of litigation resulting from the misuse or abuse of I.Q. tests gives

further credence to this recommendation. The labels resulting from the misuse of I.Q. tests and the damage to the victims of these labels are evident in our public schools.

Loss of financial assistance both for programs and for individuals have resulted from public policy based on student achievement. A case in point is the Head Start Program that actually was very successful, but was judged on ill-conceived measures and improper assessment tools. Other programs have suffered budget cuts or have been eliminated on the basis of test results, when it was the test that was the failure, not the program. The same may be said for many students who have been denied financial assistance for their education based largely or solely on test results.

The Graduate Record Examination has a very poor predictive record but universities continue to use it as an exclusionary instrument. Standardized Admissions tests for entrance to college and professional schools have such questionable predictive validity for certain segments of our student population that a moratorium would seem to benefit all applicants -- not just black applicants.

There are sufficient studies to support this position. Some of these can be found in the appended bibliography. It is both unethical and inappropriate to base public policy on any results obtained through the use of current I.Q. tests on minority groups.

2. It is recommended that a national monitoring body, with the power to enforce, through sanctions, be established to assure proper assessment and policy regarding the administration of assessment tools.

Without such a monitoring body, there is no guarantee that even legislation related to ethical testing procedures will be honored. This group should be a national body and should, by composition and sensitivity, reflect the best interests of our various minority populations along with those of the majority. Selection procedures may be determined after further deliberation.

A test or other assessment tool is no better than its user. We are concerned about what may be called the "experimenter variable," i.e., even the value of an acceptable instrument may be destroyed by one who is not qualified professionally or personally. The individual who assumes the responsibility of interpreting the test results of minority group individuals must be sensitive to the nuances of the many cultures within our pluralistic society. An insensitive person can contaminate test results or adversely affect the individual being tested by even unconscious manifestations of his insensitivity. This position, too, is based on research in this area.

3. Companies that develop, publish, and sell tests must assume (or continue to assume) a major responsibility for assuring the corrections of ills related to their product. This Task Force has strong feelings about this issue, and concurs with the other Task Forces on the mandatory nature of this obligation.

4. The appointment to public office or the nomination and election to office (particularly as that office involves the establishment of public policy based upon the testing process)

should always be based upon an individual's knowledge of, sensitivity to, and ability to converse with all segments of society, since all will be affected.

5. No individual currently in public office should make decisions about public policy based on the results of testing or research on minority groups without the concurrence of groups, organizations or individuals most knowledgeable of and conversant with life-styles, value-attitudes and "experience-in-America" of those minority groups.

ON A FAIR TESTING CODE

PREAMBLE

In a purported meritocratic and adversarial milieu in which so much credence is placed upon education, and in all of the values and status derived from it, the need for evaluative criteria and assessment tools is prominent. From the cradle to the grave there has been an increasing reliance in this country on the use of an assortment of standardized tests to screen, select, admit, reject; to classify, stratify, track, license or certify. Such subjective dependence on "objective" tools makes it possible, and even attractive in some quarters, to use them inequitably in ways which preclude a segment of the population from reaching the heights of its aspirations and realizing its potential, or from acquiring the skills and financial resources to develop their own.

In order to prevent the further misuse of tests or other measuring devices, to dispel existing myths about their infallibility, to ensure a fair and equitable use across all segments of the population and to promote the intended best use of standardized tests, the testing industry and those who subscribe to their services must become more introspective about the impact of the testing phenomenon.

Further, the testing industry must assume a greater responsibility in correcting the abuses and misuses that result from application of their products (tests), which attempt to measure intelligence, ability, aptitude, achievement, and other potentials, matters which are critically important for participating in the mainstream of society. The Task Force on the Code for Tests and Testing calls upon both the industry and the user, not only to engage in introspection, but to adopt the code which follows, and to be governed by it as they continue to work with tests and test-related activities.

CONSTRUCTION OF TESTS

Tests, regardless of their claims of objectivity, are a reflection of the experiences, characteristics, values and training of those individuals who construct them. Many of these tests are then used in the decision-making process which determines the future of minority people* who are excluded from the developmental process. Those tests which are administered widely to minority and majority people alike, and which attempt to measure intelligence, achievement, aptitude and to predict the potential of all test-takers, must employ trained persons who have experienced life as a member of a minority. Minority representation must be involved in the process of conceptualizing and developing tests. More specifically, minority persons must be involved in the overall development from its initial conceptualization to the final product. As an alternative, minorities may be forced to find financial resources and to develop their own testing programs.

* Minority people as used in this Code designates the following: Black, Spanish speaking (Puerto Ricans, Hispanics, Chicanos, Latinos); Native Americans (Indians); Asian-Americans (Japanese, Chinese, other orientals).

STANDARDIZATION PROCESS

Since most test scores are interpretable only in relationship to the group on which the test was normed or standardized, and since the norming process is so critical to the entire concept of standardized tests, the testing industry is called upon to define and make public in prominent, clear and appropriate literature the process by which its tests were standardized. Further, since there exists an assumption that blacks and other minorities are usually at or below the norm of their white counterparts, it is imperative that the norms reflect the pluralistic characteristics of the different ethnic groups that make up the tested population. Not only should minority people be included in the norming population, but the testing industry must identify the sample characteristics on and by which the test was standardized.

CONDITIONS OF ADMINISTRATION

Inasmuch as the test results of individuals frequently have immeasurable influence on their status in life, the testing industry and those who administer tests should ensure that optimal and uniform conditions always prevail. For example, where centers are established for local administration of nationwide tests consideration should be given to logistical problems, e.g., the distance candidates are required to travel. Buildings should be properly placarded so that students who are unfamiliar with the location of testing rooms, rest rooms, and other critical areas are not disadvantaged. Testing rooms should be well lighted and ventilated and should contain writing surfaces which are comfortable for all. Proctors should be hired to reflect the ethnic make up of the candidates being tested. All Proctors must be sensitive to the needs, questions and/or anxieties of all candidates.

No individual may be permitted to administer tests to members of minority groups or to interpret such data unless (a) he is duly qualified and proficient in the technical aspects of the testing process and (b) he can demonstrate a keen sensitivity to the life-styles, value-attitudes and "experience-in-America" of the several populations being submitted to the assessment procedures.

In summary, there should be a regulatory mechanism which not only monitors test administration, but one which takes the necessary corrective actions to proscribe irregular and unfair administration of tests.

LIMITATIONS OF USES

Failure to adhere to a code of conduct or to regulate oneself, on the part of the test developer and the test user could not only escalate the now rampant misconceptions that exist about standardized tests, but also invite external regulation of the testing industry. The testing industry, accompanied by the test user, must take the lead in divesting itself of the misconceptions about what tests can and cannot do, and then in dispelling similar misunderstandings among the public-at-large, and ultimately in promoting a more diagnostic, cautious and creative use of tests in the educational process, to wit:

- a. Intelligence Tests (or ability tests) -- there is a notion held by many test users that intelligence or aptitude is synonymous with an immutable or fixed characteristic within an individual. This fixation, or one's "native ability," is said to determine what is expected of one, and also one's level of expectation for all time. The testing industry must describe and publicize the fact that its intelligence or aptitude tests do not measure, in an interpretable manner, one's level of expectation throughout life. The industry must be more forthright in calling attention to the fact that learning depends not only on inherited abilities but importantly, also upon life experience in a particular environment. Since what is learned may differ according to one's economic status in life, the tests should not be used as a predeterminer of the level to which an individual aspires and may obtain.
- b. Admissions tests -- myriad tests are produced and used in the admissions process to post-secondary and to graduate institutions. Because of the nature of the admissions process, many of these tests are used not to admit but to exclude. The testing industry must define the proper use of admissions tests, and explicitly state the conditions under which tests should not be used. The extent to which a test is intended to predict one's performance at a given level dictates the extent to which the instrument must be validated on the entire entering population. The testing industry is obliged to encourage strongly such validation of admissions tests and to perfect a model or mechanism for effecting it.
- c. Occupational and Professional Tests -- tests for occupational and professional certification or entry should be job-related. A major issue in the test-selection process is whether the test measures abilities appropriate to performance in the job sought. The lack of correspondence between test requirements and job requirements invalidates the test (e.g., using broad aptitude or achievement tests for hiring firemen or policemen). Fitting the test to performance in the job has become the current objective of "criterion-referenced tests". In a job selection situation this approach makes sense. Its utilization in other assessment contexts needs further exploration.

The U.S. Supreme Court in Griggs v. Duke Power Co. held, "If an employment practice which operates to exclude Negroes cannot be shown to be related to job performance, the practice is

prohibited." Griggs v. Duke Power Co. and Chance v. New York have indicated that there must be a match (bonafide relatedness) between the test content and the skills and knowledge needed in performing a job. Where there is a mismatch between test content and expected performance, a serious barrier is established for the person who is subjected to the tests.

The Task Force on the Code for Testing calls on the testing industry to refrain from establishing contractual relationships which require variance from guidelines suggested by the courts.

CLEAR INTERPRETATION

Since misconceptions of tests can lead to damaging uses, the testing industry is obligated to make full disclosure of the purposes for which tests are designed, the processes by which they were designed, the population on which they were standardized, the statistical characteristics which delimit their use, e.g., standard error of measurement, standard error of estimate, and other "do's and don'ts" which will affect optimal use and interpretation. More specifically it is incumbent upon the testing industry to give clear and precise interpretation of the scores their tests yield. The public must be informed not only that the test scores are fallible and that their reliability is imperfect, but must be told the extent of that fallability and imperfection. The public must be informed that test scores are only a sample of a student's performance and are never more than an estimate of truth.

CONSTRUCTIVE USES OF TESTS

Effective uses can be made of tests, regardless to their type, only if the user knows what the test contains, what its purposes are, and what its limitations are. We move on the assumption that an educational system, and particularly a school within the system, implicitly guarantees that students to whom achievement examinations (tests) are administered have been taught in ways such that they can reasonably be expected to have learned the information required by the examination. Therefore, we call on the testing industry to assist school systems and other users in understanding better the content and constraints of the examinations, and in helping them understand how to make optimal use of the results. Moreover, we call upon the educational community and the public and private sectors to ensure that tests are not used or relied upon where such guarantees do not exist.

RESEARCH

We recognize that many people question whether any tests have credibility. With regard to the validity of some tests the evidence is inconclusive, even among the strongest advocates of testing. Recognizing that some device is going to be used to determine access to institutions, jobs, professions, and other opportunities within society, and to determine the various ways in which human resources are used, we assert that any testing program which results

in a significantly disproportionate distribution of scores by ethnicity must meet the most stringent validity requirements. Therefore, continued test use must be contingent upon intensified and continued research on the effects (both negative and positive) of tests, on the educational opportunities and related problems of blacks and other minorities.

Some of this research needs to be reactive in the sense that it addresses things that have occurred, and some should look ahead to what ought to be done to circumvent certain problems. Quality of the research, credibility of the researcher and methodology continue to be essential elements. The need exists for more than the traditional methodology which has characterized the testing industry's efforts in the past, and, which, for an increased number of Human Resources agencies, has also become standard practice. Just as important as these components is the way in which data from that research are interpreted. We call on the testing industry, educational systems, and public and private sectors who have a vested interest in education, to hire and use black and other minority researchers who can assist in collecting, analyzing and interpreting that critical mass of data which can help us understand the effects of tests on the problems of minorities.

BETTER TRAINING OF USERS IN INTERPRETATION

Teachers, counselors and admissions officers are publics important to the testing industry since they sit at the entrance gates through which many test takers must go if they are to realize their goals. The extent to which these publics can correctly interpret and effectively use those instruments is the extent to which decisions will be made that are fair to minority test takers. The testing industry is called upon to help bridge that gap which exists between effective use and lack of understanding, by conducting workshops and institutes, and writing special publications that are aimed at interpretation of tests.

ACTION AGENDA

The recommendations for an Action Agenda are ordered according to the publics toward whom our recommendations are directed, i.e., the testing industry, the Association of Black Psychologists, the NAACP and laymen.

THE TESTING INDUSTRY

1. Develop and publish standards of competence for those who administer, score and/or interpret tests.
2. It is known that the technical information about the test varies from one cultural group to another, often in highly significant proportion. It is essential that the testing industry provide technical information appropriate for ethnic groups for whom the testing is done.
3. Establish and fund an independent research and development corporation to identify the critical problems in assessment as they relate to minority groups; sponsor investigative research and development work involving researchers who have the endorsement of minority group professional and community associations.
4. State with clarity on all descriptive information concerning tests they publish, the specific uses for which the test is designed, the specific limitations of the instrument and a full explanation as to how the results should be interpreted.
5. Establish a national monitoring body, with the power to enforce, through sanctions, to assure proper assessment and policy of assessment tools.

ASSOCIATION OF BLACK PSYCHOLOGISTS

1. Design a project to identify some of the major tests now operating to screen individuals out of educational and employment opportunities and develop a position statement on those tests, applying the Standards for Educational and Psychological Tests, developed by a Joint Committee of the American Psychological Association, Inc., American Educational Research Association and the National Council on Measurement in Education. The results of the Project to be published and disseminated widely - to users and to clients, e.g., major school districts, counselors, and the black helping professions such as social workers and nurses.
2. Develop an empirical demonstration project for refuting the kind of conclusions drawn from I.Q. and achievement tests.
3. Design and conduct workshops around the Task Force reports including effects of race of examiner, validation and reliability, improper use of I.Q. as dependent variables in research projects, biases in test construction, problem of misinterpretation and the development of alternative means of assessment.

NAACP

1. Develop a statement regarding the rights of clients, including whether an individual

has to take an I.Q. test; where they can go to get information regarding the validity of the test.

2. Initiate legal action to develop a body of case law on the testing issue.

3. Approach the American Psychological Association regarding the inclusion in the accreditation criteria, the evaluation of the capability of the psychology training program to produce special competence in cross-cultural assessment; standards for approving programs of training and curriculum offerings in the field of psychology.

4. Establish a National Task Force to develop specific guidelines for the participation by laymen in the effort to devise satisfactory assessment procedures and disseminate the guidelines on the broadest possible basis.

5. Develop a fact sheet or "Know Your Rights" pamphlet for parents and students regarding standardized testing.

6. Urge legislation at the state and federal level establishing, outside the educational bureaucracy, a properly staffed Office for Consumer Affairs for Testing and Student Evaluation.

7. Bring together a coalition of organizations to implement the recommendations of the Task Forces.

LAYMEN

1. Know your rights as a consumer. Urge the NAACP, the Association of Black Psychologists to speak to community groups regarding the rights of individuals participating in testing programs.

2. Urge your legislators at the State and National level to sponsor legislation to establish, outside the educational bureaucracy, a properly staffed office of Consumer Affairs for Testing and Student Evaluation.

SELECTED REFERENCES

- Advisory Committee on Evaluation, in Goal, a publication of the Joint Committee on Educational Goals and Evaluations, Assembly, P.O. Box 83, State Capitol, Sacramento, Calif. 95814, November 2, 1974.
- Ainsworth, M.D., Sensory motor development of Ganda infants. In The Competent Infant, New York: Basic Books, (1973), 503-14.
- Allen, V.F., What does a Reading Test, Test? Philadelphia: Temple University College of Education. "TTT" Project Monograph Series, No. 1, USOE Contract No. OEG-0-70-2046 (721), (1974).
- American Psychological Association, American Educational Research Association, and National Council on Measurement in Education (Joint Committee). "Technical Recommendations for Psychological Test and Diagnostic Techniques." Psychological Bulletin, 51, no. 2, Part II (1964).
- American Psychological Association, American Educational Research Association, and National Council on Measurement in Education (Joint Committee). Standards for Educational and Psychological Tests, (1974).
- Anastasi, Anne, Psychological Testing. 3rd ed. New York: The MacMillan Co. (1968).
- _____. Psychological tests: uses and abuses. Teachers College Record, 62, (1961), 389-393.
- Assessing minority group children, Journal of School Psychology (Special Issue), Vol II, No. 4, (1973).
- Baratz, S.S., and Baratz, J.C., Negro ghetto children in urban education. In Linguistic-Cultural Differences in American Education. The Florida Foreign Language Reporter, 1, (1969).
- Bay Area Association of Black Psychologists. Position statement on use of IQ and ability tests. In Black Psychology, ed. R. L. Jones, New York: Harper & Row, (1972), 92-4.
- Berger, P.L. and Luckmann, P., Social Construction of Reality: A Treatise in Sociology of Knowledge, Garden City, N.Y.: Doubleday and Co. (1966).
- Black Advisory Task Force to the Child Development Associates Consortium. Collaborative Assessment: A Position. Washington, D.C.: The Child Development Associate Consortium (1974).
- Black Intelligence Test of Cultural Homogeneity (BITCH), Author, Robert L. Williams, Washington University, St. Louis, Missouri (1972).
- Brenner, A. and Stott, L.H., School Readiness Factor Analyzed, Detroit, Mich.: The Merrill-Palmer Institute (1968).
- Brown, R., How shall a thing be called. In Children with Learning Problems, edited by S.G. Sapir and A.C. Nitzburg, New York: Brunner/Mazel (1973), 378-387.
- Campbell, J.T., Tests are valid for minority groups. Public Personnel Management, 2, (1973), 70-73.
- Clark, K.B. and Plotkin, L., The Negro Student at Integrated Colleges. New York: National Scholarship Service and Fund for Negro Students (1964).
- _____. A review of the issues and literature of cultural deprivation theory. In The Educationally Deprived ed. K.B. Clark, et al. New York: Metropolitan Applied Research Center (1972), 74-92.
- Cleary, Anne T., Test bias: prediction of grades of Negro and white students in integrated colleges. Journal of Educational Measurement, 5, (Summer 1968), 115-124.
- Cohen, R., The influence of conceptual rule-sets on measures of learning ability. Race and Intelligence. Washington, D.C.: American Anthropological Association (1971), 41-57.
- Cole, M., Gay, J., Glick, J.A. and Sharp, D.W., The Cultural Context of Learning and Thinking: An Exploration in Experimental Anthropology. New York: Basic Books (1971).
- Cole, N.S., Bias in selection. Research Report No. 51. Iowa City, Ia: American College Testing Program (1972).

- Dove Counterbalanced General Intelligence Test. Author, Adrian Dove (ca. 1966).
- Deutsch, M., Katz, I., & Jensen, A.R., Social Class, Race, and Psychological Development. New York: Holt, Rinehart & Winston (1968).
- Farr, R. and Griffin, M., Measurement gaps in teacher education. Journal of Research and Development in Education, 7 (1973), 19-28.
- Flaugher, R. L., Some points of confusion in discussing the testing of black students. ETS Research Memo, RM 73-75, Princeton, N.J. (March, 1973).
- _____. The new definition of test fairness and selection; development and implications, rm Educational Testing Service, RM 7317, Princeton, N.J. (September, 1973).
- Franklin, A.J., The testing dilemma for minorities, Teachers College, Columbia University. (ERIC Document Reproduction Service No. ED 103557) (1974).
- Garcia, John, IQ: The conspiracy. Psychology Today, (Sept., 1972), 40-43, 92-97.
- Gardner, John W., Excellence, Harper & Row: New York (1971).
- Ginsberg, H., The Myth of the Deprived Child: Poor Children's Intellect and Education. Englewood Cliffs, N.J.: Prentice-Hall (1972).
- Glasser, B.G. and Strauss, A.L., The Discovery of Grounded Theory: Strategies for Qualitative Research. Chicago: Aldine (1967).
- Goodwin, Leonard & Joseph Tu., The social psychological basis for public acceptance of the social security system: The role for social research in public policy formation, American Psychologist, Vol 30, No. 9, (Sept., 1975), 875-883.
- Griaule, M., Conversations with Ogotemml: An Introduction to Dogon Religious Ideas, New York International African Institute -- Oxford University Press (1965).
- Harris, T. George, An introduction: I.Q. abuse, Psychology Today, (Sept. 1972), 39.
- Hilliard, A.G., The strengths and weaknesses of cognitive tests for young children, in One Child Indivisible, ed. J.D. Andrews, Washington, D.C., The National Association for the Education of Young Children (1975).
- Hobbs, N., ed. Issues in The Classification of Children: A Source Book on Categories, Labels and Their Consequences. San Francisco: Jossey Bass (1974).
- Hofstadter, Richard, Anti-Intellectualism in American Life, Alfred A. Knopf: New York (1974).
- Jones, R., Black Psychology, New York: Harper and Row (1972).
- Jones, R.L., ed. Labelling children culturally deprived and culturally disadvantaged. In Black Psychology, edited by R.L. Jones. New York: Harper and Row (1972), 285-294.
- Jorgensen, C., Racism in mental testing: The use of I.Q. tests to mis-label black children. In Explorations in Psychology edited by A.A. Harrison, Monterey, Calif.; Brooks/Cole (1974).
- Kamin, L.J., The Science and Politics of I.Q. Potomac, Md.: Lawrence Erlbaum Associates (1974).
- Kirkpatrick, J.J., Ewen, R.B., Barrett, R.S., and Katzell, R.A. Testing and Fair Employment: Fairness and Validity of Personnel Tests for Different Ethnic Groups. New York: New York University Press (1968).
- Klienfeld, J.S., Intellectual Strengths in Culturally Different Groups: An Eskimo Illustration. Review of Educational Research, 43, (1973), 341-60.
- Labov, W., The logic of non-standard English. In Language and Poverty, Chicago: Markham, (1970), 153-189.
- Lesser, G.S., Gordon, F., and Clark, D.G., Mental abilities of children from different social class and cultural groups. Monographs of the Society for Research in Child Development, 30 (4, Serial No. 102), (1965).

- Levi-Strauss, C. The Savage Mind. Chicago: University of Chicago Press (1966).
- Levenson, E.A., The Fallacy of Understanding. New York: Basic Books (1972).
- Lewis, D. Anthropology and colonialism. Current Anthropology 14, (1973).
- Linn, R.L., Fair test use in selection. Review of Educational Research, 43, (1973), 139-161.
- Mercer, Jane R., I.Q.: The lethal label, Psychology Today, (Sept., 1972), 44-47, 95-97.
- _____. Sociocultural factors in the education evaluation of Black and Chicano children; Paper presented at the 10th Annual Conference on Civil and Human Rights of Educators and Students, Washington, D.C.: (Feb. 18-20, 1972).
- Milholland, J.E., Culture-Fair Intelligence Test. In Sixth Mental Measurement Yearbook, Tests and Reviews: Intelligence Group, ed. O. Buros. Highland Park, N.J.: Gryphon Press (1965).
- Moore, T.E., Cognitive Development and the Acquisition of Language. New York: Academic Press (1973).
- Myrdal, G., Objectivity and Social Research. New York: Pantheon Books, (1967).
- Nobles, W., Psychological research and the black self-concept; A critical review, J. Social Issues 29, 1, (1973), 11-31.
- Rippey, R.M., ed. Studies and Transactional Evaluations. Berkeley, California: McCutchan (1973).
- Roberts, S.O. and Shane, M.D., Selecting superior students for Fisk. Fisk News, 36 (No. 3 Spring), (1962), 5-6.
- _____. and Birch, J.S., College careers of superior Negro American high school graduates. Report to Carnegie Corporation of New York, from Fisk University (Department of Psychology) (1966).
- _____. and Copenheim, D.B., The Effect of Special Instruction upon Test Performance of High School Students in Tennessee. Princeton, N.J.: Educational Testing Service (1966).
- _____. and Horton, C.P. & Roberts, E.T., SAT vs GRE performance of Negro American college students. Proceedings, 77th Annual Convention, APA, (1969), 177-178.
- _____. and Horton, C.P., The Relationship of Selected Non-Intellectual Factors to Freshman Academic Achievement. Unpublished manuscript, Fisk University (Department of Psychology), (1970).
- Ryan, W., Blaming the Victim. New York: Vantage Press (1971).
- Samuda, R.J., Psychological Testing of American Minorities. New York: Dodd, Mead & Co. (1975).
- Senna, C., The Fallacy of I.Q. New York: The Third Press (1973).
- Stanley, J.C., Predicting college success of the educationally disadvantaged. Science, 171, (1971), 640-647.
- Stone, Charles C., Testing and the educational power struggle, Integrated Education: Race and Schools, Vol IX, No. 4, (July-August, 1971), 4-10.
- Terman, L.M. and Merrill, M.A., Stanford-Binet Intelligence Scale: Manual, Etc. Boston: Houghton Mifflin (1960).
- The Journal of Afro-American Issues, Vol III, No. 1, Winter, 1975, (Issue on Testing, Measurements and Afro-Americans).
- The 1975 Statement of the American Psychological Association on Testing and Minorities and replies to the same.
- Cleary, T.A., Humphreys, L.G., Kendrick, S.A. and Wesman, H.A. Educational uses of tests with disadvantaged students; American Psychologist, 30, (1975), 15-41.

- Jackson, G.D., Comment on the Report of the Ad hoc Committee on Educational uses of tests with disadvantaged students: Another psychological view from the Association of Black Psychologists. American Psychologist, 30, (1975), 88-93.
- Bernal, E.M., Comment - On the Report of the Ad hoc Committee on Educational uses of tests with disadvantaged students: A response to "Educational Uses of Tests with Disadvantaged Subjects": American Psychologist, 30, (1975), 95-96.
- Humphreys, L.G., Addendum - To the Report of the Ad hoc Committee on Educational uses of tests with disadvantaged students: American Psychologist, 30, (1975), 95-96.
- Cronbach, L.J., Five decades of public controversy over mental testing. American Psychologist, 30, (1975), 1-14.
- Thomas, A. and S. Sillen, Racism and Psychiatry, New York: Brunner/Mazel (1972).
- Thorndike, R.L., Concepts of culture-fairness. Journal of Educational Measurement, 8 (1971), 63-70.
- Torrance, E.P., Gifted Children in the Classroom. New York: The Macmillan Company (1965).
- Turner, L.D., Africanisms in the Gullah Dialect, New York: Arno Press, New York Times (1969), originally published 1949, University of Chicago.
- Watson, Peter, I.Q.: The Racial Gap, Psychology Today, (Sept., 1972), 48-50, 97-99.
- Webb, E.J., Campbell, D.T., Schwartz, R.D., and Sechrest, L., Unobstrusive Measures: Non-Reactive Research in the Social Sciences. Chicago: Rand McNally and Company (1966).
- Weber, D.A., Racial bias and the LSAT: A new approach to the defuse of present admissions, Buffalo Law Review (1975), 439-462.
- Wechsler, D., Manual for the Wechsler Adult Intelligence Scale. New York: Psychological Corporation, (1955).
- Williams, E.B., Test Bias and Predictive Validity. Paper presented at the Annual Meeting of the American Psychological Association, Chicago (1975).
- Williams, R.L., Abuses and misuses in testing black children. In Black Psychology, edited by R.L. Jones, New York: Harper and Row (1972), 77-91.
- _____, BITCH gives Blacks a break on tests. Norfolk Journal and Guide. Saturday, August 23, 1973, 11.
- _____, Black pride, academic relevance, and individual achievement. In Crucial Issues in Testing, edited by R.W. Tyler and R. M. Wolf. Berkeley, Calif.: McCutchan (1974), 13-20.

Test Bulletins

- College Entrance Examination Board (CEEB), ATP-Student Bulletin, 1975-76. Princeton, N.J., 1975-76.
- Educational Testing Service (ETS), Graduate Record Examination (GRE) Information Bulletin 1975-76. Princeton, N.J., 1975-76.
- American College Testing Program (ACT), Medical College Admission Test (MCAT) Bulletin, Iowa City, Iowa, 1975-76.
- Educational Testing Service - Law School Admissions (LSAT) Bulletin. Princeton, N.J., 1975-76.
- Psychological Corporation. Bulletin of Information - List of test centers - (MAT) Miller Analogies Test and other tests, New York, 1974.
- American College Testing (ACT), Seminars for ACT users. Iowa City, Iowa: ACT, Author, 1975.
- Educational Testing Service (ETS), Programs of Continuing Education. Princeton, N.J.: ETS, Author, 1975.